Computers and Mathematics with Applications 66 (2013) 1971–1980

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

A Multi-Objective Evolutionary Algorithm for enhancing Bayesian Networks hybrid-based modeling

Carlos R. García-Alonso [a], Pilar Campoy-Muñoz [b,*], Melania Salazar Ordoñez [b]

[a] Universidad LOYOLA Andalucía, Departement of Quantitative Methods, Campus of Córdoba, C/ Escritor Castilla Aguayo 4, 14004, Córdoba, Spain
[b] Universidad LOYOLA Andalucía, Department of Economics, Campus of Córdoba, C/ Escritor Castilla Aguayo 4, 14004, Córdoba, Spain

ARTICLE INFO

Keywords:
Multi-Objective Evolutionary Algorithms
Fuzzy inference
Bayesian networks
Monte-Carlo simulation
Remittances

ABSTRACT

Bayesian Networks are increasingly being used to model complex socio-economic systems by expert knowledge elicitation even when data is scarce or does not exist. In this paper, a Multi-Objective Evolutionary Algorithm (MOEA) is presented for assessing the parameters (input relevance/weights) of fuzzy dependence relationships in a Bayesian Network (BN). The MOEA was designed to include a hybrid model that combines Monte-Carlo simulation and fuzzy inference. The MOEA-based prototype assesses the input weights of fuzzy dependence relationships by learning from available output data. In socio-economic systems, the determination of how a specific input variable affects the expected results can be critical and it is still one of the most important challenges in Bayesian modeling. The MOEA was checked by estimating the migrant stock as a relevant variable in a BN model for forecasting remittances. For a specific year, results showed similar input weights than those given by economists but it is very computationally demanding. The proposed hybrid-approach is an efficient procedure to estimate output values in BN.

1. Introduction

Remittances have been defined as "transfers made by migrants who are employed and have lived, at least one year, in other economies" [1]. In the case of developing countries, where the number of migrants by 2010 exceeds 171 million people (3% of the population), remittances have increased significantly over the past two decades, from US$60,000 MM in 1990 to US$325,000 MM in 2010 [2]. These financial flows amount to 2% of the GDP for developing countries and in some cases outnumber 25% of the GDP [3]. Remittance size, its counter-cyclical nature and their resilience have turned them into the second largest source of foreign currency for the developing world. Consequently, the sustainability debt of those countries is stronger, making their access to international capital markets also easier [3]. However, empirical studies have showed that remittances can also have negative effects on labor supply, inflation and real exchange rate, and mixed effects on the economic growth of the recipient economies [4].

In recipient economies, the estimation of the macroeconomic effects of remittances is critical. In this context, there have been some attempts to estimate remittance flows by using econometric techniques [5,6]. These studies highlighted that the lack of information and the poor quality of available data are the main obstacles for accurate forecasting. In addition, the economic literature about remittances reveals the complex nature of this phenomenon, in which a plethora of macroeconomic and microeconomic variables are involved [7]. On this basis, García-Alonso et al. [8] dealt with a remittance estimation problem by combining Bayesian Networks (BN), Monte Carlo simulation (MCS) and Fuzzy logic (FL).

* Corresponding author. Tel.: +34 957 222 100; fax: +34 957 222 181.
E-mail address: mpcampoy@uloyola.es (P. Campoy-Muñoz).

0898-1221/$ – see front matter
doi:10.1016/j.camwa.2013.01.029

Expert-knowledge elicitation and learning from data can be used for stochastic assessment of BN. The former is a widely used alternative to develop BN when data is scarce or does not exist, but this process may be costly and time-consuming [9]. When data is available, several approaches can be used to develop BN, including structural equations [10] and learning algorithms [11]. Structural equations need prior expert knowledge to define the basic structure of the BN and then this method calculates the relevance of variables and their relationships according to a strong set of initial hypotheses [12]. Learning algorithms are devoted to discovering both the BN structure [13–15] and its parameters [16–18] from data. However, these algorithms can be very sensitive to the initial setting chosen by researchers and large amount of data are required [19]. In addition, given the complexity and stochastic nature of complex systems, algorithms have to operate in huge search spaces and they could be easily trapped into the numerous sub-optimal solutions [19]. To overcome some of these drawbacks, meta-heuristic approaches have been applied, such as evolutionary algorithms [20,21], evolutionary programming [22,23] and ant colony optimization [24,25]. Finally hybrid-approaches, combining expert knowledge and automated learning, have also being designed and developed to guide and speed up the process of learning efficient BN structures and parameters [26,27].

In this paper, the approach developed by [8] is enhanced by including a Multi-Objective Evolutionary Algorithm (MOEA). In [8], the relative importance (weight) of each input variable in its BN fuzzy dependence relationship (DR) is always defined by the experts as a random variable. The aim of this paper is to estimate these weights automatically when DR output values are known (secondary information). The determination of these weights can be seen as an Optimization Problem (OP), in which the combination of weights that optimizes a predefined fitness function could be found by using a bio-inspired searching procedure. Given an expert-based BN structure, including both probabilistic and fuzzy DR, the MOEA calculates input weights for the fuzzy ones knowing all or some output values. Thus, MOEA contributes to the evaluation of fuzzy DR and lets effective forecasting by learning the relative input relevance. Therefore, the hybrid-based approach offers a new strategy for stochastic evaluation of BN by combining the expert knowledge, Monte-Carlo simulation and Fuzzy inference with bio-inspired learning from data procedures.

This paper is organized as follows. Section 2 briefly describes the hybrid-based approach for stochastic assessment of BN, focusing on the role of input variable weights in fuzzy DR. Section 3 explains the OP regarding to input weight determination and offers a detailed description of the MOEA. In Section 4, the complete hybrid-based approach is used to estimate the migrant stock, as a relevant variable in the remittances model, and empirical results from a real case are shown. Finally, Section 5 concludes and discusses the future work.

2. The methodological framework

2.1. Bayesian Networks

Bayesian Networks (BN) have rapidly become a leading tool for modeling complex systems under uncertainty by combining graph theory and probability theory [10]. BNs are graphical representations of conditional dependence relationships among stochastic variables. In a BN each node is a stochastic variable and when a causal relationship between two nodes exists it is represented by an arrow. A BN can have a very complex structure including bidirectional arrows and cycles. In a directed acyclic graph G, that is a special type of BN, each node represents a random variable $X = [X_1, \ldots, X_q]$ in D_X and arcs encode direct conditional dependence relationships between variables $X_h \to X_i$, where X_h is the parent of X_i and, in turn, X_i is the descendant of X_h. In G no cycles are permitted. Given the parents of the discrete variable X_i, denoted by pa_i, its conditional distribution is defined by $P(X_i = x_i) = p(x_i|pa_i)$. Thus, the joint probability distribution of G is:

$$p(x_1, x_1, \ldots, x_n) = \prod_i p(x_i|pa_i). \tag{1}$$

Thus, a BN can fully describe qualitative and quantitative aspects of the real-world problem under study. Given the very frequent lack of data on socio-economic systems, BNs have to be developed mainly through expert knowledge. In these cases, experts and/or selected literature are consulted about the structure of the BN: nodes and arcs. According to [10], the structure of a graph G can be considered a BN when (i) it includes all the relevant variables and (ii) their relationships have a causal nature (Markovian compatibility). The stochastic evaluation of a BN can be intractable when dealing with complex systems because of the increasing number of variables and their relationships. However, given the BN structure, Monte-Carlo simulation and fuzzy logic can be considered an alternative to assess BN in complex socio-economic systems [8].

2.2. The dependence relationship structure

A BN can have two types of Dependence Relationships (DR): the probabilistic DR (denoted by f) and the fuzzy DR (denoted by u). Both DRs relate input variables (parental nodes) – always random variables-X_i with their corresponding output variables (descendant nodes) – consequently, random variables-Y describing the causal behavior of the system under study. A probabilistic DR $Y = f(X_i)$ usually has an algebraic structure that can be easily managed using, for example, Monte-Carlo simulation. On the other hand, a fuzzy DR $Y = u(X_i)$ is a set R of fuzzy "*IF*…*THEN*" standard rules ($d = 1, 2, \ldots, c$):

$$R : \underset{d=1}{\overset{c}{\text{ALSO}}} \left[\text{IF} \underset{i=1}{\overset{n}{\text{AND}}} (x_i \in X_i \; isr \; A_{id}) \; \text{THEN} \; y_d \in Y_d \; isr \; B_d \right] \tag{2}$$